Title

Single-Vehicle Fatal Crash Prediction for Two-Lane Rural Highways in the Southeastern United States

Authors
Zhu, Hong
Dixon, Karen K.
Washington, Simon
et al.

Publication Date
2010

Peer reviewed

SafeTREC Safe Transportation Research & Education Center

Single-Vehicle Fatal Crash Prediction for Two-Lane Rural Highways in the Southeastern United States

Hong Zhu and Karen K. Dixon, Oregon State University,
Simon Washington, SafeTREC, and David M. Jared,
Georgia Department of Transportation

2010
RR-2010-10 http://www.safetrec.berkeley.edu

Single-Vehicle Fatal Crash Prediction for Two-Lane Rural Highways in the Southeastern United States

Updated November 15, 2009

Authors:

Hong Zhu
Oregon State University
School of Civil and Construction Engineering
220 Owen Hall
Corvallis, OR 97331
Phone: 541-740-3216
Email: zhuh@onid.orst.edu

Karen K. Dixon, Ph.D., P.E. (corresponding author)
Oregon State University
School of Civil and Construction Engineering
220 Owen Hall
Corvallis, OR 97331
Phone: 541-737-6337
Fax: 541-737-3052
Email: karen.dixon@oregonstate.edu

Simon Washington, Ph.D.
UC Berkeley Traffic Safety Center
2614 Dwight Way #7374, Berkeley, CA 94720-7374
Phone: 510-642-0566
Fax: 510-643-9922
Email: tscenter@berkeley.edu

David M. Jared, P.E.
Georgia Department of Transportation
Office of Materials and Research
15 Kennedy Drive
Forest Park, GA 30297-2534
Phone: 404-363-7569
Fax: 404-363-7684
Email: djared@dot.ga.gov

Abstract: 227
Total world count: 5238 words + [(6 tables + 3 figures) x 250 words each] = 7488 equiv. words

ABSTRACT

The rural two-lane highway in the Southeastern United States is frequently associated with a disproportionate number of serious and fatal crashes and as such remains a focus of considerable safety research. The Georgia Department of Transportation spearheaded a regional fatal crash analysis to identify various safety performances on two-lane rural highways and offer guidance for identifying suitable countermeasures to mitigate fatal crashes. The fatal crash data used in this study were compiled from Alabama, Georgia, Mississippi, and South Carolina. The database, developed for an earlier study, included a total of 557 randomly selected fatal crashes from the years 1997 and/or 1998 (varied per state). Each participating state identified the candidate crashes and performed physical or video site visits to construct crash databases with enhance site-specific information.

Motivated by the hypothesis that single- and multiple-vehicle crashes arise under fundamentally different circumstances, the research team applied binary logit models to predict the probability that a fatal crash is a single-vehicle run-off-road fatal crash given roadway design characteristics, roadside environment features, and traffic conditions proximal to the crash site. A wide variety of factors appears to influence or be associated with single-vehicle fatal crashes. This paper also includes a model transferability assessment where the authors determined that lane width, horizontal curvature, and ambient lighting are the only three significant variables consistent for the single-vehicle run-off-road crashes for all study locations.

INTRODUCTION

The rural two-lane highway in the Southeastern United States is frequently associated with a disproportionate number of serious and fatal crashes and as such remains a focus of considerable safety research. The Georgia Department of Transportation (GDOT) spearheaded a regional fatal crash analysis to identify various safety performances on two-lane rural highways and offer guidance for identifying suitable countermeasures to mitigate fatal crashes. This study used physical site data from an earlier research effort to assess potential ways to address perceived safety hazards for these locations. In October 2005, the GDOT and researchers with the Georgia Institute of Technology completed a summary report that identified a series of rural two-lane road safety assessments for states in the southeastern United States. This study effort resulted in four randomly sampled similarly formatted fatal crash databases from Alabama, Georgia, Mississippi, and South Carolina and provided a unique data source so that a cross-sectional comparison study could be performed. This paper reports an assessment of single-vehicle fatal crashes that can help illuminate candidate treatments and enhance road safety for rural two-lane highways.

Prior to initiating an analysis, the research team conducted a literature review to determine what available safety models are published that could apply to this target rural road environment. Since road characteristics and the policies that establish the design of roads vary across jurisdictions, the published literature is limited to assessment of physical road features between jurisdictions and generally focuses on crashes within individual jurisdictions. Much of the published literature has been based on crashes from only one state or select study corridors, while other studies focused on national fatality data. The various studies demonstrate many contradictions about crash type, severity and associated factors justifying the need to further explore both crash conditions and the rural road environment.

This study investigates cross state differences and similarities for rural two-lane highway fatal crashes from Alabama, Georgia, Mississippi, and South Carolina in terms of the impact on crash conditions and potential contributing factors. The study's findings ultimately help to explain the various safety performances of these highways among the four states and offers guidance for generating countermeasures to specifically mitigate single-vehicle fatal crashes. The literature review also identified five primary causal influences for crashes: vehicle occupant/driver, vehicle characteristic, road and roadside features, crash characteristics, and environmental conditions.

Efficient and effective safety predictive models can vary based on specific objectives such as what to predict, at which level to predict, and which method to use. The vast majority of safety prediction models attempt to predict crash frequency (number of crashes that occurred during a period of time) or crash rate (crash frequency over the traffic exposure). Common models often used for these assessments include Poisson regression models, negative binomial regression models, or variations of these models. Systematic-level safety measures, such as number of crashes that occurred over a time period for specific road segments, require analysts to aggregate (or sort) crash counts into categories and extract road geometric and roadside information for each individual road segment. In this case, a variable only represents an average condition of the corresponding road segment rather than reflecting a unique feature of a crash site.

Safety performance prediction at an aggregated level is important for roadway network screening and facility evaluation as it can help identify problematic areas. However, these systematic models do not permit evaluation at the individual crash level. In some cases, road or crash characteristics may have unique associations with safety measures at a disaggregated level that can be different than at the category or aggregated level. This phenomenon is known as ecological fallacy, an error that occurs when falsely assuming individuals in a group have the average attributes of the group as a whole. In other cases, some crash level variables are inappropriate for aggregation.

In addition to the more common crash frequency and crash rate safety measures, a considerable number of researchers have also developed models to predict crash injury severity levels and their associated crash costs (*1,2,3*). Crash type, in contrast, has been minimally investigated (see for example *4* and 5). Some researchers have addressed crash types by predicting crash frequency for a specific type of crash alone (*6,7,8*). This method can increase homogeneity of crash data since crash records would include only one specific type of crash, such as head-on crashes. Analysis of a homogenous dataset, however, may lead to the exclusion of potential relationships across different types of crashes.

For this effort, the research team has elected to focus on crash types for fatal crash analysis at two-lane rural highways. The focused evaluation of fatal crash types may help reveal crash type associations that will be different from what can be determined when studying all crashes. Secondly, certain crash types tend to be over-represented for a specific type of highway facility. For example, a single-vehicle run-off-road crash type makes up about 60% of overall fatal crashes on two-lane rural highways; while the single-vehicle crash is less common and generally less severe at urban locations.

Fatal crash type prediction models can serve as an analytical assessment tool for safety improvement projects where the main goal is to reduce fatalities and severe injuries. In the process of identifying locations with the greatest safety needs, most current methods do not directly consider specific crash types. Crash type prediction models provide an approach to quantify safety performance measures by taking into account roadway design characteristics, road environmental features, as well as traffic conditions. A random sample of fatal crash records is available for this analysis. These records include detailed crash-specific roadway geometric characteristics that were collected through site investigation or video log inspection. This comprehensive data set enabled the authors team to better understand the relationships between roadway design features, fatal crashes, and single versus multiple vehicle crash occurrence.

REGIONAL FATAL CRASH DATA

As previously indicated, researchers from several states in the southeastern part of the United States previously participated in a rural fatal crash study. This paper includes a cross-sectional analysis of data from four of these states. Alabama and Georgia researchers randomly selected 150 two-lane rural highway fatal crashes for the year 1997 by assigning random numbers to all fatal crashes for that year in the Fatality Analysis Reporting System database (FARS) and then selecting a sample based on simple number selection. Though conceptually more than one crash could have occurred at the same road segment, the resulting data set did not happen to include multiple crashes at adjacent locations (*9*). The FARS database contains all fatal traffic crashes in the United States including those that occur in the 50 states, the District of Columbia, and Puerto Rico. For a crash to be included in the FARS database, all resulting fatalities of vehicle occupants and non-motorists must have occurred within 30 days of the crash. Mississippi had a smaller crash population and so researchers from that state similarly developed a random sample of 100 fatal crashes for the year 1997. South Carolina evaluated 157 fatal crashes in their final analysis. These South Carolina fatal crashes occurred during 1998. Additional information regarding the random sampling procedures can be found in the state-specific project reports (*10, 11, 12*).

Each participating state identified the candidate crashes and performed physical or video site visits within two to three years of the crash. The data collection process included identification of unique physical features not commonly included in crash reports. Where available, the research team acquired this information via video analysis; however, non-state maintained roads as well as many of the state roads required physical site visits to acquire information such as lane width, roadside ratings, shoulder treatment, or similar. The only variable that could not be confidently identified using the video approach was pavement edge drop off, so this was not included as a variable in the models. The physical road infrastructure information merged with the crash data makes this dataset one of the largest available datasets of its kind. Each state used the same fatal crash site data collection form so that the data from each site and state would be consistent (*10*). Generally, each database included five types of data elements: crash details, site characteristics, environmental factors, limited driver information, and vehicle characteristics. All fatal crashes on rural two-lane highways were included in the study; however, this paper only focuses on the run-off-road single-vehicle crashes that made up a significant portion of the study database.

METHODOLOGY

For this study, the research team's goal was to develop models to predict fatal crash type outcomes. Since crash type is a categorical variable, the research team performed categorical data analysis with a logistic regression model to predict categorical response variables with both continuous and categorical predictors. This study defines crash types based on the definition of "Manner of Collision" in the "First Harmful Events" category used by FARS (*13*). Figure 1 presents the fatal crash type classification structure. FARS describes the first harmful event as either the first property damage or injury-producing event of a crash occurrence. The single-vehicle run-off-road crash identifier is applied when the first harmful event is a non-collision (e.g. driving off a cliff, rollover), a collision with an object that is not fixed (e.g. pedestrians or animals), or a collision with a fixed-object (e.g. trees, utility poles). Since there are only a few crashes striking objects that are not fixed, a simplified representative description for this study is a single-vehicle crash where the vehicle exited the roadway and either struck a fixed object or overturned. For crashes that involved more than one vehicle, the two major fatal crash types observed for this data set were head-on collisions and angle crashes.

After initial examination, the research team developed two types of crash type prediction models as follows:

- Single-vehicle fatal crash vs. Multiple-vehicle fatal crash
    - Based on fatal crash history for rural two-lane highways, predict the probability of a single-vehicle fatal crash.
- Head-on fatal crash vs. Other fatal crash (not a head-on)
    - Based on multiple-vehicle fatal crash history for rural two-lane highways, predict the probability of a head-on fatal crash.

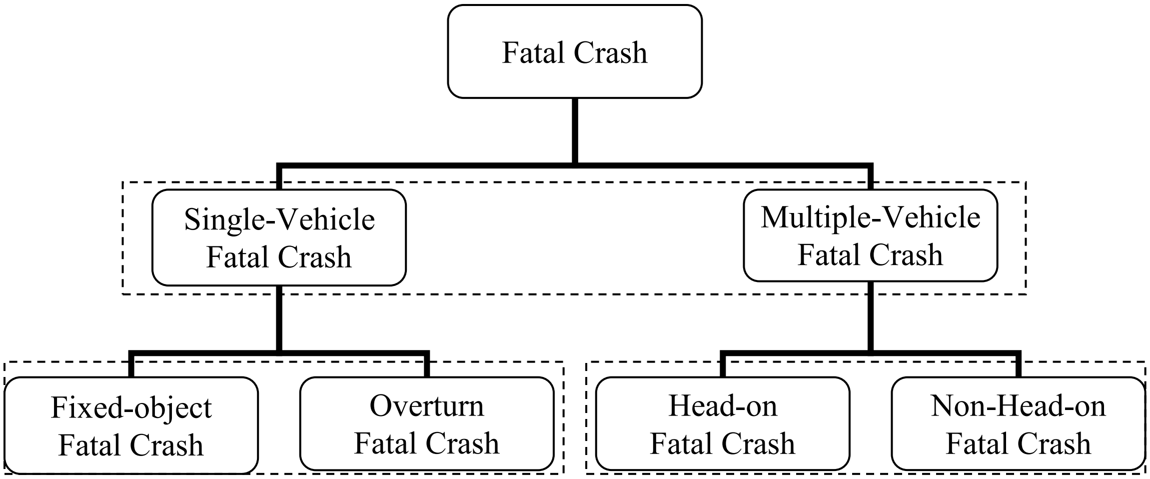

Figure 1: Fatal Crash Type Classification

This paper only presents the crash type prediction models that can differentiate a single-vehicle fatal crash from a multiple-vehicle fatal crash. Further information of the head-on fatal crash versus other fatal crash models can be found in the final project report (*14*).

As previously discussed, one of the objectives of this study was to develop crash type models based on the premise that roadway design characteristics, roadside environment features, and traffic conditions each directly influence crash type occurrence. The research team applied a binary logit model to help identify influential factors that can differentiate two crash types at a time.

To compare conditions using a binary logit model, the first step is to assign a value of either zero or one for the crash type of interest. For this effort, the research team assigned the following values for variable Y for each fatal crash:

Y= 1, if the crash type is a single-vehicle run-off-road fatal crash;

Y= 0, otherwise.

The analyst can then use the binary logit model to estimate the probability that Y has the value of 1 based on independent variables that represent features associated with the crash conditions (X_1 ,….., X_k). The logistic function form estimates what the probability would be of observing a single-vehicle run-off-road crash in the event of a fatal crash, as shown below:

$$\Pr(Y=1)=\Pr(Single-veh-runoff)=\frac{\exp(\beta_0+\sum_{i=1}^{i=k}\beta_i X_i)}{1+\exp(\beta_0+\sum_{i=1}^{i=k}\beta_i X_i)} \quad (1)$$

Given:

Pr(Single-veh-runoff): the probability of observing a single-vehicle run-off-road fatal crash occurrence in the event a fatal crash occurred, this will be a value between 0 and 1;

β_0: estimated intercept;
β_i: estimated coefficient for the corresponding independent variable X_i;
X_i: the i^{th} independent variable.

Common model verification procedures frequently used for ordinary regression models and some types of logistic regression models are not suitable for evaluating a binary logit model (*15, 16*). The model evaluation and selection for binary logit models mainly relies on examining the significance of extra terms in the model including squared terms or possible interactions between variables. For this analysis, influential variables that appeared to help significantly differentiate crash types for the final model (with a p-value less than 0.10) were retained during the examination of potential contributing factors.

The ultimate goal for developing safety predictive models was to identify valuable information and quantify relationships between highway design characteristics and associated safety performance. While the statistical significance and model goodness-of-fit are very important considerations in this process, the authors also made decisions based on t heir knowledge of how highway design characteristics relate to crash types. This known relationship between design characteristics and safety performance explains why the ultimate final models may include higher-than-typical p-values of 0.1 or larger. Though a common p-value for statistical analysis is 0.05, the authors maintained the 0.1 value so as to demonstrate the performance of select variables that slightly exceed the conventional value. This relaxed p-value permits analysis of a wider variety of variables that could potentially have a less significant influence on the crash condition.

RESULTS AND DISCUSSION

The research team developed fatal crash type prediction models to estimate the probability of a single-vehicle run-off-road fatal crash occurrence in the event of a fatal crash, based on the four-state combined fatal crash database (AL, GA, MS, SC), three-state combined database (AL, GA, SC), and state specific databases. Variables used in the state-combined models are defined in Table 1. Model development contrasted crash information from each state with similar crashes in Georgia (the “base state”). For the four-state model, the estimation results were not significant for state indicator variable AL and SC, but the MS p-value was significant. This observation indicates that fatal crash types in Georgia, Alabama, and South Carolina share similar characteristics compared to the disparate findings for similar crash types in Mississippi. In other words, the predictor variables in the models for GA, AL, and SC are sufficient to explain crash differences across these states whereas they are not for MS. Since one objective of this study was to identify rural two-lane highway fatal crash models that can help analysts better understand crash trends in Georgia and the other states, the research team also investigated the combined-state model based on the fatal crash database from AL, GA, and SC (the three similar states).

Table 1: Variable Description (Combined-State Models, Single-Vehicle)

Types	Variables	Descriptions
Location Indicator	AL	1 if in Alabama, 0 otherwise
	MS	1 if in Mississippi, 0 otherwise
	SC	1 if in South Carolina, 0 otherwise
Road Junction type	JUNCTION	1 if a road junction, 0 if road segment
Geometric design features	LW	Lane width (ft)
	PSW	Paved shoulder width (ft)
	GSW	Graded shoulder width (ft)
	LCURV	1 if curve to the left, 0 otherwise (curve to the right or straight alignment)
	CREST	1 if vertical crest curve, 0 otherwise
Roadside condition	RHR67*	1 if road hazard rating is 6 or 7 (hazardous and not traversable), 0 otherwise
Traffic volume	ADT	Average daily traffic (10^3 veh/day)
Land use type	LU_C	1 if commercial driveways in the proximity (500 ft longitudinally of crash site) of the crash location, 0 otherwise
Lighting condition	DARKUNLIT	1 if dark with no supplemental street lights, 0 otherwise
Circadian biological clock	HR_DEEPSLEEP	1 if crash time between 1a.m. and 3a.m., 0 otherwise
	HR_EM	1 if crash occurred between 12a.m. – 6 a.m., 0 otherwise
Safety Protection	RESTRAINT	1 if driver wore safety restraint, 0 otherwise

*Note: The RHR is defined by the seven scale roadside hazard rating system which was developed to characterize the potential crash risk through roadside design features based on a pictorial scale on two-lane highways (*17*).

In the three-state model, the state indicator variables AL and SC again did not yield significant estimation results. These results imply that the three-state model does a suitable job of predicting crashes for Georgia-specific conditions.

The combined three-state model benefits from a larger sample size, with approximately 428 fatal crashes. This larger sample size enabled the authors to investigate more potential contributing factors at various significance levels. Meanwhile, the research team also developed state-specific models in order to investigate the opportunity of examining spatial transferability as well as to explore unique variables that may only have impacts on fatal crash outcomes in one or two states.

Three-State Model (AL, GA, SC)

Table 2 summarizes descriptive statistics for the continuous and categorical variables used for analysis in the multi-state models. Table 4 presents the resulting three-state and Georgia model estimations and their goodness-of-fit test results. The resulting three-state single-vehicle run-off-road fatal crash prediction model as depicted in Table 3 is presented in equation format as:

For:

$$\eta_{3-state} = 6.6717 - 0.1855AL - 0.1167SC - 0.8078JUNCTION - 0.5407LW - 0.0542PSW - 0.0475GSW$$
$$-0.0676(PSW * GSW) + 0.788LCURV - 1.7264CREST + 2.5199(LCURV * CREST) + 1.1581RHR67 - 0.0965ADT$$
$$-1.3722LU\_C + 1.3101DARKUNLIT + 1.8318HR\_DEEPSLEEP$$

The probability of a single-vehicle run-off-road fatal crash can be predicted for a given set of road and environment conditions as:

$$\Pr(Single - veh - runoff)_{3-state} = \frac{\exp(\eta_{3-state})}{1 + \exp(\eta_{3-state})} \qquad (2)$$

Table 2: Distribution of Continuous and Categorical Variables (Three-State Model, Single-Vehicle)

Summary of Continuous Variables				
Variable	**Mean**	**Std Dev**	**Minimum**	**Maximum**
LW (ft)	10.8	1.1	8	12
PSW (ft)	0.6	1.6	0	12
GSW (ft)	5.2	3.5	0	16
ADT (veh/day)	2,896	2,941	75	17,960

Summary of Categorical Variables		
Variable	**Status**	**Percent (%)**
JUNCTION	0 (Segment)	75
	1 (Intersection)	25
LCURV	0 (Curve to Right or Straight)	75
	1 (Curve to Left)	25
CREST	0 (Not a Crest Vertical Curve)	89
	1 (Crest Vertical Curve)	11
RHR67	0 (Roadside Hazard Rating < 6)	92
	1 (Roadside Hazard Rating of 6 or 7)	8
LU_C	0 (No Commercial Driveways)	94
	1 (Near Commercial Driveways)	6
DARKUNLIT	0 (Daylight, Dark with Lights, Dusk, or Dawn)	55
	1 (Dark without Supplemental Lights)	45
HR_DEEPSLEEP	0 (Not between 1 a.m. and 3 a.m.)	96
	1 (From 1 a.m. until 3 a.m.)	4

Table 3: Single-Vehicle Fatal Crash Model Estimation (Three-State and Georgia Models)

Variable	Three State Model (AL, GA, SC)		Georgia Model	
	Estimate	P- value	Estimate	P- value
Intercept	6.6717	<.0001	8.9011	0.0003
AL	-0.1855	0.5614	-- [2]	-- [2]
SC	-0.1167	0.7404	--	--
JUNCTION	-0.8078	0.0051	-2.1473	<.0001
LW	-0.5407	0.0003	-0.8350	0.0003
PSW	-0.0542	0.5873	-0.3506	0.0647
GSW	-0.0475	0.3202	--	--
PSW*GSW [1]	-0.0676	0.0929	--	--
LCURV	0.7880	0.0156	1.7437	0.0112
STRAIGHT	--	--	1.5662	0.0046
CREST	-1.7264	0.0002	--	--
LCURV*CREST [1]	2.5199	0.0223	--	--
RHR67	1.1581	0.0716	--	--
ADT	-0.0965	0.0558	--	--
LU_C	-1.3722	0.0313	--	--
DARKUNLIT	1.3101	<.0001	1.1195	0.0124
RESTRAINT	--	--	-1.1604	0.0151
HR_DEEPSLEEP	1.8318	0.0926	--	--

	Three State Model (AL, GA, SC)			Georgia Model		
Observations (Single-Vehicle/Others)	428 (259/169)			146 (85/61)		
AIC	440.976			151.893		
SC	505.922			175.762		
-2 Log L	408.976			135.893		
R-Square	0.3204			0.3484		
	Hosmer and Lemeshow Goodness-of-Fit Test			Hosmer and Lemeshow Goodness-of-Fit Test		
	Chi-Square	DF	Pr > Chi-Square	Chi-Square	DF	Pr > Chi-Square
	7.4889	8	0.4849	4.983	7	0.662

[1] Note: LCURV*CREST and PSW*GSW indicate these variable pairs interact.
[2] Note: The corresponding variable is not included in that model.

Among the 428 fatal crashes available for the three-state combined model, 259 crashes were single-vehicle run-off-road fatal crashes. As presented in Table 4, the Hosmer and Lemeshow Goodness-of-Fit test showed an acceptable goodness of fit (p-value = 0.4849 >0.05) for the three-state model. As shown in Equation (2), variables that can significantly differentiate single-vehicle fatal crash from multiple-vehicle fatal crash include the presence of a road intersection (junction), lane width, paved shoulder width, graded shoulder width, horizontal curve direction, presence of a crest vertical curve, a roadside hazard rating 6 or 7, average daily traffic, driveway land use type, lighting condition, and time of crash. This model indicates, for example, that in the event a fatal crash occurs, the probability of a single-vehicle run-off-road fatal crash will increase at road segments with horizontal curves to the left. Single-vehicle run-off-road fatal crashes tend to occur less frequently as both the lane width and traffic volume increase.

This study examined the potential correlation among predictors. It demonstrated that the explanatory variables do not appear to be strongly correlated with others. Therefore the potential multi-colinearity is of minimal concern for model development. The selection of potential predictors targeted road design features and attributes from roadside environment conditions that can potentially be addressed using engineering countermeasures. Though there are indications that alcohol and drug abuse are a primary contributor to the vehicle occupant/driver influence, these speculations cannot be quantitatively confirmed due to known quality limitations for this specific data variable. Additionally, variables that appear to be strong predictors for crash frequency prediction may not be significant factors in the fatal crash type model. For example, crash time of day is often recognized as a significant factor for crash frequency prediction as this variable might inadvertently represent traffic volume fluctuations, driver fatigue, or lighting conditions. For single-vehicle crashes, however, crash time does not significantly differentiate single-vehicle fatal crashes from multiple-vehicle fatal crashes. This may be because more than 60-percent of the fatal crashes occurred during night time conditions. The variable called HR_DEEPSLEEP is intended to capture the influence of the human circadian cycle while at the same time this variable might also provide other indications about driver condition and possible impairment. In addition, the authors treated some variables as dichotomous variables rather than maintaining their original scales. For example, the roadside hazard rating, RHR, is a variable with seven scales. Only the road side conditions that pose high crash risks (RHR of 6 or 7) are likely to be associated with single-vehicle fatal crashes. Therefore it is reasonable to re-categorize RHR as a dichotomous variable in the single-vehicle fatal crash type model.

Up to 80-percent of the fatal crash locations in the 3-state crash database did not include paved shoulders. The overall shoulder width (paved and graded shoulder width) provides an approximate representation of the effect from graded shoulders. The potential safety effects from paved shoulders could easily be masked if included with a general shoulder variable. Therefore, the authors approached the shoulder related variables by treating paved and graded shoulder width individually and examined the potential interaction effects in order to identify more informative results. The statistically significant interaction effect between the paved and graded shoulder width indicates that the influence of the paved shoulder width on the probability of single-vehicle fatal crashes is also dependent on the graded shoulder width. Similarly, the horizontal curve to the left and the presence of a crest vertical curve exhibit a similar interaction effect.

This study also illustrates how sensitive the probability or a single-vehicle crash is based on varying graded shoulder widths (0, 2, 4, 6, and 8 ft) at various levels of paved shoulder width. The probabilities can be calculated based on a set of pre-determined base conditions for a typical study road segment crash, see Table 4. Most of the variables were assigned a value similar to their average condition in the sample data with a crash location in Georgia (GA=1). As shown in Figure 2, and Figure 3, the probability of a single-vehicle fatal crash when the graded shoulder width is increased does not vary substantially if there is no companion paved shoulder (paved shoulder width = 0 ft). Alternatively if there is a paved shoulder present, the probability of a single-vehicle fatal crash drops significantly when the graded shoulder width is increased. This relationship suggests that the combination of paved and graded shoulders collectively helps to enhance safety and reduce the likelihood of single-vehicle fatal crashes.

For daylight conditions or locations with supplemental lighting, the probability of a single-vehicle fatal crash occurring decreases at a slower rate for wider graded shoulders than when the lighting conditions are dark with no supplemental lighting. This observation is particularly true at locations with a wider paved shoulder.

Table 4: Description of Basic Condition for Evaluating Single-Vehicle Models

Variables	Conditions
AL	0
SC	0
JUNCTION	0 (a road segment)
LW	11 ft (lane width = 11 ft)
LCURV	0 (road horizontal alignment is not a curve to the left)
CREST	0 (road vertical alignment is not a crest vertical curve)
RHR67	0 (roadside hazard rating 1 through 5)
ADT	3000 veh/day (average daily traffic estimated as 3000 veh/day)
LU_C	0 (not in the proximity of a commercial driveway)
HR_DEEPSLEEP	0 (crash did not occurred between 1 a.m. – 3 a.m.)

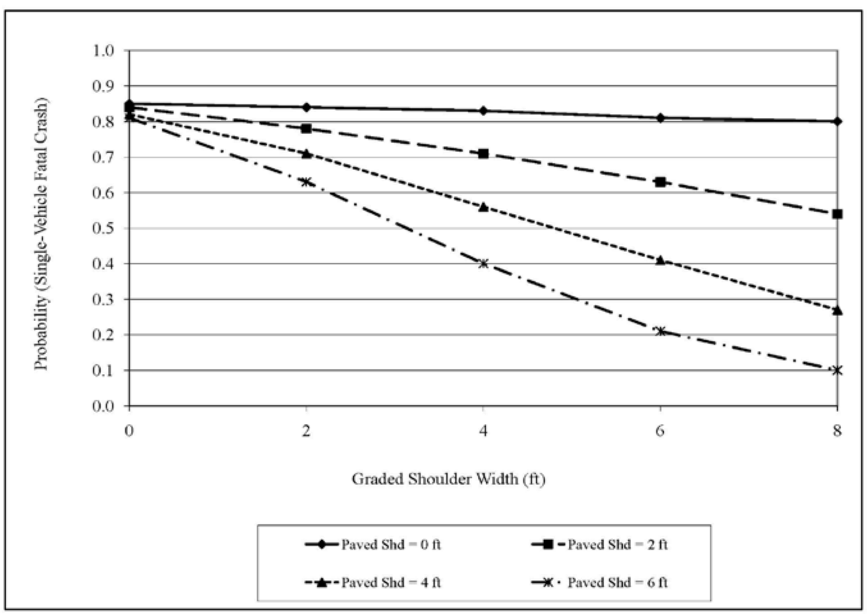

Figure 2: Dark without Street Lights -- Graded and Paved Shoulder Width (Three-State Model, Single-Vehicle)

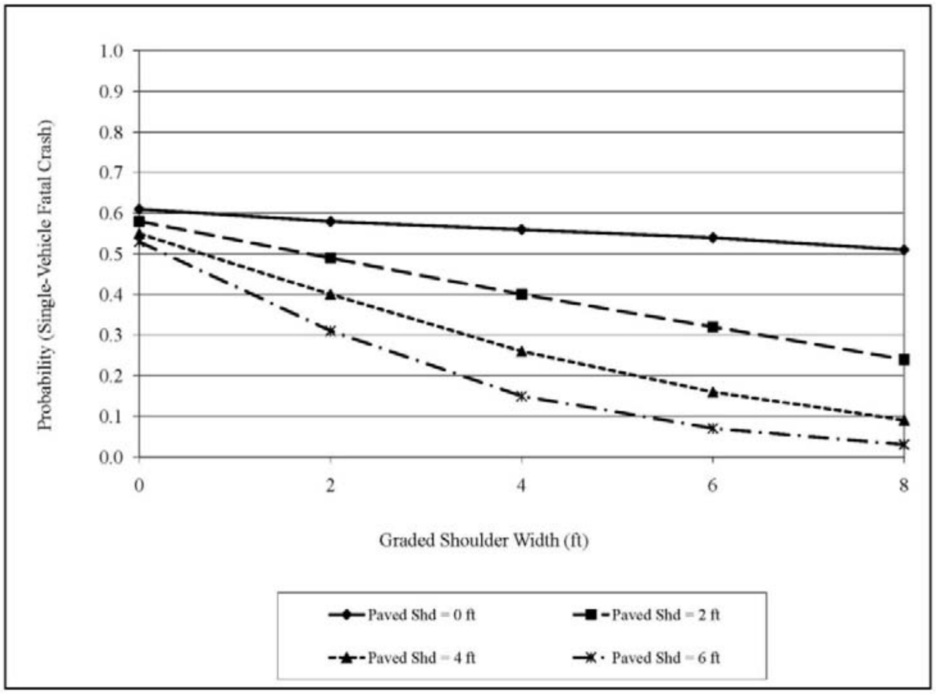

Figure 3: Daylight, Dark with Lights, Dusk, or Dawn -- Graded and Paved Shoulder Width (Three-State Model, Single-Vehicle)

GA Only Model

Four individual models, based on the fatal crash sample data, were estimated for each of the states of Alabama, Georgia, Mississippi, and South Carolina. This paper presents the Georgia-only model in detail; however, the summary discussion addresses all four separate state models. In addition to the modeling effort for a regional level as summarized by the three-state combined model (AL, GA, SC), the individual state assessment can be used as an indicator for identifying potential state-specific significant influential factors and their corresponding effects on t he probability of single-vehicle run-off-road fatal crash occurrence. This state-level modeling effort can also help determine the suitability of model transferability for other state applications.

Independent variables which have significant impacts on the fatal crash type outcomes in the Georgia-only model include intersection (junction) type, lane width, paved shoulder width, horizontal curve direction, horizontal alignment type, roadside lighting condition, and safety restraint system usage for at-fault drivers. In the Georgia fatal crash database, lane widths ranged from 8 ft to 12 ft with average lane width at 10.7 ft (approximately 11 ft). Average width of paved shoulders was 0.6 ft but for a total range from 0 ft to 6 ft width.

The resulting Georgia-only single-vehicle run-off-road fatal crash prediction model as depicted in Table 4 is presented as follows:

Let:

$$\eta_{GA} = 8.9011 - 2.1473JUNCTION - 0.835LW - 0.3506PSW + 1.7437LCURV + 1.5662STRAIGHT$$
$$+1.1195DARKUNLIT - 1.1604RESTRAINT$$

The probability of a single-vehicle run-off-road fatal crash for Georgia can then be predicted as follows:

$$\Pr(Single-veh-runoff)_{GA} = \frac{\exp(\eta_{GA})}{1+\exp(\eta_{GA})} \quad (3)$$

As shown in Table 4, 85 out of 146 fatal crashes were single-vehicle run-off-road fatal crashes. For the state of Georgia, if a fatal crash occurred, the likelihood of a single-vehicle run-off-road fatal crash would increase at a location with a horizontal curve to the left after accounting for other factors. Other significant variables for the Georgia single-vehicle fatal crash include roadside hazard ratings of 6 or 7, dark without supplemental lighting, and at-fault drivers not utilizing safety restraints. There were about 71-percent at-fault drivers who did not wear safety restraints at the time of crash. Meanwhile, single-vehicle fatal crashes are less likely to occur at locations with increasing lane and paved shoulder widths. The maximum lane width for the Georgia highway crash sites was 12 ft and the maximum paved shoulder width was 6 ft so this observation should not be extrapolated outside these upper boundaries.

Model Discussion

Table 5 summarizes the model estimation results for the four individual-state models, the three-state combined model (AL, GA, SC), and the four-state combined model (AL, GA, MS, SC). The four individual-state models do not contain the same set of independent variables—suggesting that a multi-state model represents a compromise model that captures many of the primary but possibly not all factors associated with fatal crashes. The two combined-state models include a collection of independent variables included in all four individual-state models, except for few variables. The effort of fitting four individual-state models with the same set of independent variables was not supported by the data—again highlighting the previous point. One of the requirements of testing model spatial transferability is to fit models with the same set of predictors. This condition can only be achieved if all four individual-state models include a limited collection of independent variables such as ADT only. These models would then have less accurate predictive power since critical contributing factors may be excluded.

Both of the combined state models provided similar estimates for the categorical location indicator variables for Alabama and South Carolina indicating that the fatal crash type outcome prediction is substantively similar across at least three states: Alabama, South Carolina, and the base state Georgia.

The two combined-state models present very similar modeling results when the same set of independent variables is retained. As shown in Table 5, despite the differences among the individual-state models and the combined-state models, there are three independent variables (lane width, horizontal curve direction, and lighting conditions) that are significant predictors with similar effects for all six models. Rural roadway segments with narrower lanes have a greater likelihood of single-vehicle run-off-road fatal crashes than their wider lane counterparts. The location with a curve to the left tends to be more frequently associated with a single-vehicle run-off-road crash than are locations with either curves to the right or straight alignment. Similarly, the location that is dark without supplemental lighting is more likely to be a site for a single-vehicle run-off-road crash than locations with better lighting conditions or daytime conditions. These findings suggest that the lane width, curve direction, and lighting condition are strongly associated with the probability of a fatal crash type, with direct associations for single-vehicle fatal crashes.

Table 5: Model Comparison (Single-Vehicle)

Variables	**AL only Model**	**GA only Model**	**MS only Model**	**SC only Model**	**Three-State Model (AL, GA, SC)**	**Four-State Model (AL, GA, MS, SC)**
AL					-0.1855	-0.1984
MS						-1.3453**
SC					-0.1167	-0.0836
JUNCTION	-1.2158**	-2.1473**			-0.8078**	-0.9922**
LW	-0.5111**	-0.835**	-0.4282*	-1.0576**	-0.5407**	-0.4630**
PSW		-0.3506*	-0.7045		-0.0542	-0.1087
GSW				-0.1247*	-0.0475	-0.0463
PSW*GSW					-0.0676*	-0.0622*
LCURV	1.5906**	1.7437**	0.8562	1.1555**	0.7880**	0.7255**
STRAIGHT		1.5662**				
CREST				-1.5341**	-1.7264**	-1.5389**
DOWN[a]	0.8342*					
LCURV*VCREST					2.5199**	2.2686**
RHR67	1.8195		1.6633**		1.1581*	1.3314**
ADT					-0.0965*	-0.1078**
LU_C				-2.5984**	-1.3722**	-1.4298**
DARKUNLIT	1.3682**	1.1195**	1.6260**	1.3523**	1.3101**	1.3135**
HR_DEEPSLEEP			2.0741*	2.0821	1.8318*	1.9744**
HR_EM[b]	2.8888**					
RESTRAINT		-1.1604**				

** Significant level < 0.05
* Significant level < 0.1
[a] Note: 1 if direction of slope is down (negative), 0 otherwise.
[b] Note: 1 if crash occurred between 12a.m. – 6 a.m., 0 otherwise.

Practical Application

Safety engineers can apply fatal crash type prediction models as a unique tool for safety improvement projects and can use the models to specifically focus on reducing fatalities and serious injuries. Since current assessment techniques do not always incorporate crash types, the use of predictive models can complement current procedures to identify candidate countermeasure applications. It is helpful for safety engineers to know whether a candidate improvement location tends to have higher likelihood of a major fatal crash type based on the existing road design characteristics. This assessment can occur on newer roads that do not have a substantial crash history if these roads are built in a manner consistent with others in the region.

Assume that a high-crash location has been previously identified using regional analysis procedures and that the road is a rural two-lane highway. This two-lane rural road segment has a known history of single-vehicle crashes that result in fatalities or serious injury. One specific location on this road has the existing characteristics as shown in Table 6 along with the estimated results for the probability of a single-vehicle fatal crash at the example study location based on the existing conditions, and the two proposed improvement conditions, B1 and B2.

Table 6: Sample Problem -- Existing Road Conditions for Georgia Site

Existing Condition	**Status**	**Variables**
Road segment?	Yes	JUNCTION = 0
Alabama?	No	AL = 0
South Carolina?	No	SC = 0
Lane width	11 ft	LW=11
Paved shoulder width	0 ft	PSW = 0
Graded shoulder width	8 ft	GSW = 8
Roadside hazard rating	5	RHR67 = 0
ADT	3,000 vehicles per day	ADT = 3
Land use	Driveways not for commercial use	LU_C = 0
Driving during 1am to 3am?	No	HR_DEEPSLEEP = 0
Curve to the left?	Yes	LCURV = 1
Crest?	No	CREST = 0
Daylight, dark with lighting, dusk or dawn conditions	-	DARKUNLIT = 0
Dark without supplemental street lights	-	DARKUNLIT = 1

Evaluation Results

	Lane Width (ft)	Paved Shoulder Width (ft)	Graded Shoulder Width (ft)	Probability of Single-vehicle fatal crash	
				Daylight	Dark No Lighting
Existing	11	0	8	0.70	0.90
Plan: B1	12	0	8	0.57	0.83
Plan: B2	11	3	5	0.45	0.75

<u>Evaluation conclusion:</u>

- Proposed plan B2, shoulder improvement, is the recommended countermeasure under the context of single-vehicle fatal crash outcome reduction.
- Since the physical improvements have less influence on single-vehicle crashes during dark conditions, it may be appropriate to enhance the location (particularly at horizontal

curve locations) with other countermeasures that specifically increase safety during dark conditions.
- Due to the increased risk due to the lack of safety restraint use by at-fault drivers, it is recommended that the use of safety restraints be promoted.
- Though the probability models provide indications regarding the effectiveness of improvements, the final improvement decisions should be based on cost/benefit analysis, as well as other potential conditions not available for assessment in the model development.

Application Limitation

The correct use of logit model results can be problematic if it is not clear how the models should be used and what limitations should be applied for use of the models. This study focuses on fatal crashes where at least one person was fatally injured. It is not appropriate to generalize the modeling results to crashes at all injury levels. The models developed for this study include a limited number of contributing factors. There are other potential factors that could influence fatal crashes, but these variables are not included in the model due to a variety of reasons. For example, the random fatal crash database may have some variables that are not well populated and therefore do not provide significant effects. It is also possible that there may be influential variables that are not available in the standard crash database or the supplemental database used for this study.

CONCLUSIONS

The use of statistical models to predict how a candidate countermeasure can help to reduce a specific type of crash can be valuable. The research team evaluated a wide variety of statistical models and determined that a logit model is a suitable tool for determining the probability of a crash and, by doing so, would help determine how to reduce the crash probability. Many options are available for estimating crash severity, frequency, and type. This paper specifically evaluates crash type. Since rural two-lane roads have a high number of single-vehicle crashes, the crash type evaluated extensively in this research was a single-vehicle run-off-road crash. For the single-vehicle crashes, the following observations were identified:

- Single-vehicle fatal crashes in Mississippi did not have similar contributing factors than similar crashes in the other three states, so the cross-section model for single-vehicle crashes only applies to Alabama, Georgia, and South Carolina. This is likely due to a Mississippi data quality issue.
- There are a wide variety of variables that influence a single-vehicle fatal crash in the three states. These include location, lane width, shoulder width and type, horizontal curve direction, crest vertical curves present, horizontal and vertical geometric interactions, roadside hazard rating, traffic volume, driveway type, lighting conditions, and crash time.
- Individual single-vehicle models for the four states have similar influences, and consistently critical influences for fatal crashes in all four states include lane width, horizontal curve direction, and lighting conditions.
- For the Georgia-only model, the use of safety restraints and lighting conditions were critical factors associated with single-vehicle fatal crashes.

Potential countermeasures that can be evaluated using the single-vehicle fatal crash type model are categorized as follows:

- Geometric alignment improvements;
- Widening of lanes or pavement widths;
- Adding or widening graded or stabilized shoulder; and
- Widening or improvement of clear zones.

These countermeasure categories are also recommended in a previous countermeasure evaluation study that focused on fatal crashes at rural roads in Georgia. For that project, the countermeasure recommendations were based on expert panel opinions (*14*). The independent recommendations from the previous study closely align with the roadside hazard rating, the graded/paved shoulder condition, and the horizontal and vertical curve variables (each contributing to crashes as determined through the use of the statistical models). The corroboration of these two independent assessment techniques solidifies the creditability of both the expert opinions and the statistical models.

ACKNOWLEDGEMENTS

This research was sponsored by Georgia Department of Transportation and the authors would like to acknowledge their support for this effort. In addition, the authors would like to thank university researchers and representatives of the Departments of Transportation in the States of Alabama, Mississippi, and North Carolina.

REFERENCES

1. Abdel-Aty, M.A. Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models. *Journal of Safety Research*, Vol.34, No.5, 2003, pp. 597-603.

2. Shankar, V., F. Mannering, and W. Barfield. Statistical Analysis of Accident Severity on Rural Freeways. *Accident Analysis and Prevention*, Vol. 28(3), 1996, pp. 391-401.

3. Wang, X.K., and K.M. Kockelman. Use of Heteroscedastic Ordered Logit Model to Study Severity of Occupant Injury. In Transportation Research Record: Journal of the Transportation Research Board, No.1908, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 195-204.

4. Kim, D.G., J. Oh and S. Washington. Modeling Crash Outcomes: New Insights into the Effects of Covariates on Crashes at Rural Intersections. *ASCE Journal of Transportation Engineering*. Vol. 132 No. 4, 2006.

5. Shankar, V., Mannering, F., and W. Barfield. Effect of Roadway Geometric and Environmental Factors on Rural Freeway Accident Frequencies., *Accident Analysis and Prevention*, Vol. 27, No. 3, 1995, pp. 371 – 389.

6. Lee, J., and F. Mannering. Impact of Roadside Features on the Frequency and Severity of Run-Off-Roadway Accidents: An Empirical Analysis. *Accident Analysis and Prevention*, Vol. 34(2), 2002, pp. 149-161.

7. Abdel-Aty, M.A., and H. Abdelwahab. Modeling Rear-End Collisions Including the Role of Driver's Visibility and Light Truck Vehicles Using a Nested Logit Structure. *Accident Analysis and Prevention*, Vol. 36(3), 2004, pp. 447-456.

8. Jonsson, T., J.N. Ivan, and C. Zhang. Crash Predication Models for Intersections on Rural Multilane Highways: Differences by Collision Type. In Transportation Research Record: Journal of the Transportation Reseach Board, No. 2019, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 91-98.

9. Dixon, K. K. *Final Summary Report: Southeastern United States Fatal Crash Study*. Report No. FHWA-GA-05-9815. Georgia Department of Transportation and Federal Highway Administration, 2005.

10. Washington, S., K.K. Dixon, D. White, and C.- H. E. Wu. *Final Report: Investigation of Fatal Motor Vehicle Crashes on Two-Lane Rural Highways in Georgia.* Report No. FHWA-GA-02-9905, Georgia Department of Transportation, 2002.

11. Clarke, B.D., and W.A. Sarasua. *Final Report: Fatal Crash on Rural Secondary Highways.* Report No. FHWA-SC-03-04, South Carolina Department of Transportation, 2003.

12. Mississippi Department of Transportation. Investigation and Identification of Principal Factors Contributing to Fatal Crashes. 2000.

13. U.S. Department of Transportation and National Highway Traffic Safety Administration. *2007 FARS Coding and Validation Manual.* Web address: http://www-nrd.nhtsa.dot.gov/Pubs/FARS07CVMAN.PDF (Accessed Jan 29, 2009).

14. Dixon, K.K., H. Zhu, S. Washington, and K. Shin. *Final Report Examination of Crash Trends in the Southeastern US: Analysis of Fatal Crashes*. Report No. GDOT Project 07-01, Georgia Department of Transportation, 2009.

15. Agresti, A. *Categorical Data Analysis.* 2nd edition, John Wiley & Sons, Inc, 2002 (ISBN: 978-0-471-36093-3), 2002.

16. Ramsey, F.M., and D.W. Schafer. *The Statistical Sleuth: a Course in Methods of Data Analysis.* 2nd edition, published by Duxbury, 2002 (ISBN 0-534-38670-9).

17. Zegeer, C. V., D. W. Reinfurt, J. Hummer, L. Herf, and W. Hunter. *Safety Effects of Cross-Section Design for Two-Lane Roads. Transportation Research Record 1195*, Transportation Research Board, 1988.